DOCUMENT RESUME

ED 093 953 TM 003 771

AUTHOR Holtz, Ruth Richards
TITLE Entering Attitudes of Undergraduates in Education Toward Topics in Educational Measurement.
PUB DATE Apr 74
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, Illinois, April, 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *College Students; Course Objectives; *Education Majors; Individual Differences; *Measurement; *Student Attitudes

ABSTRACT

Three hundred and thirty-one undergraduates in education who had not taken educational measurement rated 47 course objectives for importance and interest. Median ratings were positive or neutral for all items. Importance and interest were highly related. The most favored items involved critical analysis of or alternative approaches to measurement procedures; least favored dealt with statistics. Factor-analytic results and differences between males and females, students in elementary and secondary education, and various subspecialties were also considered. Some suggest that attitude may block achievement in measurement. Results show students have definite preferences within the curriculum. However, they enter instruction with an overall positive disposition. (Author)

ED 093953
TM 003 271

ENTERING ATTITUDES OF UNDERGRADUATES IN EDUCATION TOWARD TOPICS IN EDUCATIONAL MEASUREMENT*

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Ruth Richards Holtz
Boston University

From an extensive study involving 2,877 senior education majors in 86 teacher training institutions, Mayo (1967a) concluded that beginning teachers do not possess a high degree of knowledge or skill in educational measurement, and that they show only small gains in competency after two years of classroom experience. Goslin (1967) emphasizes the importance of adequate training in measurement, stressing the large amount of standardized testing done in schools. Teacher-made tests are used even more frequently but are often poor in quality (Stanley and Hopkins, 1972). Glaser (1973) suggests that testing could become an even more integral part of the educational process in the future, with a shift to criterion-referenced tests and a continual matching of student capabilities to educational options. The problem of teacher competency in measurement is a critical one.

It has been suggested (e.g., Goslin, 1967a; Mayo, 1967, 1970a) that attitude toward educational measurement may be an obstacle to achievement in course work or to later implementation of measurement principles. Mayo (1967) proposes that many students may be hindered by (a) a negative attitude toward statistics--statistics may be seen as irrelevant to most teachers' needs--and by (b) a more general failure, in spite of course work, to appreciate the usefulness of measurement principles, with a resulting (c) conceptual and affective gap between teachers and students of measurement. Possible positive and negative consequences of testing have been noted (e.g., Bloom, 1968; Goslin, 1967b). Some students may be focusing on perceived negative implications of present measurement practice, linking measurement with a variety of assumed educational and social ills (Schutz, 1971). Indeed, many recent writings of educational reform view testing as a "central evil to be challenged and exorcised (Glaser, 1973, p. 564)." Such writings may both reflect and influence student attitude toward educational measurement and, perhaps, toward instructors of measurement.

Preliminary to constructing the test used in his survey of teachers' measurement competency, Mayo (1967) surveyed a variety of measurement specialists and educators on the importance of 70 measurement objectives; these objectives were developed from a content outline of an NCME committee. Most objectives were seen as

---

*Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois, April 15-19, 1974.

important by this sample, with those relating to classroom tests most important in general, followed by objectives involving standardized tests, uses of measurement and evaluation, and statistics. Measurement students' priorities may be somewhat different. There is a need for a similar survey of attitudes of the beginning measurement student to determine importance, and relative importance, of specific measurement objectives. Meaningfulness and relevance of less attractive material could be stressed, with possible gains in both attitude and mastery (Mayo, 1967, 1970a, 1970b). Relationships between measurement objectives perceived as less and more important might be emphasized, in particular.

Methods

For the present study, a list of 47 measurement objectives was developed. Items were chosen to relate, at a more general level, to those included in the Mayo list, and to cover possible measurement topics not included by Mayo (e.g., criterion-referenced tests, affective objectives, historical precedents of present testing practice). A few items were suggested by recurring requests of students in measurement classes and related suggestions in the literature (e.g., Mayo, 1970a, 1970b)--for example, to develop ways to minimize test anxiety or let students have input in defining classroom objectives. Most objectives involved activities at or above the application level of Bloom et al's (1956) cognitive taxonomy. Those items involving specific uses of tests were presented as evaluation level objectives and were supplemented by several affective items which involved defining one's own position on measurement alternatives; it was felt that some students might consider these items important whether positive or negative toward current measurement practice.

Items were written in non-technical language, or with technical terms defined, where included, and were presented in an order which was randomly determined. The general introductory stem, "A course in educational measurement can enable me to:" was included at the top of each page. Subjects were asked to rate each objective for importance, using a five point scale which ranged from "very important" to "very unimportant."[1] A further response option ("X") was provided for those who did not understand the statement. Students were asked to put down a number to identify their papers,[2] but were not asked to give their names and were assured of the confidentiality of their answers. They were told the researcher was interested in student attitudes toward aspects of an educational measurement course and were encouraged to respond as carefully and candidly as possible.

Subjects were 234 undergraduates in education who had not yet taken a measurement course.[3] One-hundred and eighty students were in secondary education, and 54

were in elementary education. Most secondary students were administered the instrument as the initial activity in their educational measurement course. Elementary students beginning measurement could not be included (as naive subjects) as they were already involved in another study. Instead, the instrument was given to several educational psychology sections of a required education core course, typically taken by sophomores, the year before measurement. The instrument was administered by the author to all but about 15 of the subjects.

Results and Discussion

The typical objective was understandable to almost all of the respondents. Ratings for only one item (#46, correlation coefficient) had to be interpreted cautiously, due to a high proportion of "X" responses.[4] There was considerable similarity between responses of the 27 secondary students in the educational psychology classes and the remaining secondary sample.[5] On the assumption that elementary responses would also be similar across contexts, the elementary sample was also included for comparison in the data analysis.

The sections below address: (a) overall rank ordering of objectives by importance (b) differences between males and females, and students in elementary and secondary education, and (c) factor analytic structure of objectives.

Rank Order of Objectives by Importance

Objectives ranked by median importance for the entire sample are presented in Table 1. The median was employed due to distributional skew. The Spearman correlation between this list and rankings from the smaller elementary subsample is 0.93; thus, priorities of both elementary and secondary groups are reflected in this list. Median ratings ranged from 1.19 (very important) down to 2.94 undecided). In general, no items were felt to be unimportant, and ratings were generally skewed toward positive importance. However, variation was considerable for all objectives, with the full range of attitude present in almost every case. (The typical standard deviation was about 1.0--and standard error of the mean about 0.07--for these items).

In general, items rated highest in importance involved critical analysis of or alternative approaches to measurement procedures; the least favored dealt with statistics. Items involving similar general content were not always considered similar in importance. For example, items involving objectives were primarily

high in importance, while those involving test construction and standardized tests, although lower in general, were more varied in importance. Items on reliability and validity were somewhat varied but generally high. Items addressed to uses of tests were rated high in importance in the context of feedback and goal setting, and improving curriculum and instruction, but were lower when applied to placement and grouping, and grading.[6]

Results suggest that students are: (a) positive or somewhat positive in general to the importance of most testing and measurement procedures, with perceived importance varying as a function of the types of procedures involved and their intended use, and (b) undecided, overall, as to the importance of statistics. This was a somewhat more complex ordering of priorities than that of Mayo's group of educators. Students also seemed (c) particularly concerned with evaluating measuring instruments and procedures and deciding on their own use of testing methods.

It is notable that many of the objectives perceived as most important involve higher level cognitive or affective goals. A large number relate to the synthesis and, particularly, evaluation levels of Bloom et al's (1956) cognitive taxonomy; many less important objectives are lower in taxonomy level. Interestingly, many of the lower-ranked--and, often, lower-taxonomy--objectives are useful and even required to address some of those objectives considered most important. Use of statistical methods in estimating instrument reliability and validity is one notable example. These are connections which need to be stressed in instruction.

Differences Between Groups

Elementary Females, Secondary Males and Females. While there were many more similarities than differences between subgroups of this sample, some of the differences are of interest. Table 2 includes results from significant univariate analyses of variance and Dunn multiple comparisons between means, following a significant multivariate F-ratio, for 48 elementary females, 120 secondary females, and 58 secondary males. (Subjects who didn't report their sex were eliminated as were the few males in the elementary education group.) Several significance levels need to be interpreted with caution, due to inhomogeneity of variance.[7] General results will be emphasized here.

All but two significant differences distinguished secondary males and females--and, in some cases, elementary females as well. Females were more positive than males toward each of these items, a trend which was generally borne out in the non-significant items. Typically, means for elementary and secondary females were similar, such that a larger elementary sample might have caused these variables to distinguish males and females generally, across groups. Distinguishing

items involved diverse content (test usage, item-writing, statistics). A number of the most distinguishing variables involved relatively less popular aspects of measurement.

Results suggest more differences between males and females than between elementary and secondary students generally. This may partially reflect the fact that elementary and secondary students have had essentially the same course experience in education prior to the year of measurement. Beyond this, (secondary) females may be somewhat more receptive than (secondary) males, in general, to aspects of the curriculum. (Results might also reflect the fact that a female typically administered the instrument to these subjects.)

Area of Specialization. Information on area of specialization was available on some secondary subjects. These subjects were divided into groups of males and females in humanities/social science (English, language, social studies education) and math/sciences (mathematics, science, business education), yielding 26 males and 51 females in humanities/social science (H/SS) and 15 males and 26 females in math/sciences (M/S). Most of the subjects were in English and social studies, and math and business.

Multivariate analysis of variance showed significant main effects on both sex and specialty; there was no significant interaction. Most of the sex differences in the previous analysis reappeared for this sample. The second part of Table 2 gives the significant differences between H/SS and M/S groups.

M/S students gave more importance than H/SS students to constructing good objective items and evaluating test uses in grading. H/SS students were more positive toward the remainder of the items. Greatest differences involved constructing good essay questions and recognizing situations in which norms may be useful or misleading in score interpretation (both $p < .0001$). Remaining differences involved a variety of items (testing atmosphere, standardized tests and norms, testing programs, historical aspects, statistics).

H/SS and M/S students differ as one might expect on types of item writing considered important, suggesting different instructional emphases for these groups. The more general positive disposition of H/SS may suggest relatively higher interest in understanding and making one's own decisions about selected aspects of testing. It does not necessarily suggest more positive disposition toward present measurement practice.

Factor Analysis

A factor analysis was performed on data from the entire sample. A principal components solution produced 13 factors with eigenvalues greater than one,

accounting for 62.6% of the variance. These factors were rotated to the normal varimax criterion. Results are presented in Table 3. The present data didn't lend itself readily to simple structure.[8] Beyond this, factor identity was often unclear; factors did not necessarily include items with similar content. (Some factors can, however, be tentatively identified.[9])

The present analysis indicates that relationships between some variables are complex. Items involving, for example, objectives, standardized tests, and even statistics, are somewhat dispersed among factors, with factors such as I defined by a seemingly heterogeneous group of items. This may indicate that students see similarities between some objectives as a complicated function of type of procedure, purpose, and context of use. It may alternatively indicate a lack of conceptual clarity about the curriculum. In some cases (e.g., Factor VIII), objectives of similar content load on the same factor but are notably different in perceived importance, perhaps indicating a perceived differential effectiveness in addressing the same problem. Instruction might create different relationships between measurement objectives and greater similarity in intra-factor item importance, as students see new applications and relevance(s) of various measurement procedures.

Conclusions

While there were a few differences, subgroups in this sample felt similar in general toward topics in educational measurement; overall, the present sample seemed not to be negatively disposed toward the curriculum. Students were generally undecided about the importance of statistics, and varied in attitude toward the importance of other objectives, results somewhat different than those from the Mayo (1967) sample of educators. Also, there was notable student interest in addressing certain measurement issues, and in evaluating certain measurement procedures, both in general and in the context of their own future teaching situations.

Findings are contrary to Mayo's suggestion of a general negative disposition toward measurement. While this sample could be different, Mayo's general observations would seem to apply to Boston University. Perhaps a vocal minority makes itself well heard; indeed, distribution of response was broad on most items, and some respondents were very negative on almost every item on the instrument. Or, perhaps, negative attitude is developed or increased during instruction. Mayo (1967, 1970a) suggested that the relevance of some measurement material may not be adequately communicated. The present study suggests that for the entering student, the full relevance of some measurement material to high-priority student objectives has yet to be established. Perhaps, in addition, material students feel is most

important is not always stressed in instruction, and student expectations for the course are not realized. Negative attitude could certainly increase in this event.

Instruction in measurement could focus on existing student concerns and relate material considered less important to these--always emphasizing its relevance and importance. Mayo (1970a) reports some innovative teaching techniques which appear promising. With such an emphasis, undergraduates in education might come to more fully realize the relevance of measurement principles and techniques to their future goals as teachers.

References

Bloom, B.S. (Ed.) Taxonomy of educational objectives, handbook I: cognitive domain. New York: Longmans, Green & Co., 1956.

Bloom, B.S. Toward a theory of testing which includes measurement-evaluation-assessment. Occasional Report No.9. Los Angeles: Center for the Study of Evaluation of Instructional Programs, University of California, 1968.

Glaser, R. Educational psychology and education. American Psychologist, 1973, 28(7), 557-566.

Glass, G.V., Peckham, P.D., & Sanders, J.R. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. Review of Educational Research, 1972, 42(3), 237-288.

Goslin, D.A. Teachers and testing. New York: Russell Sage Foundation, 1967a.

Goslin, D.A. The social impact of testing. Personnel and Guidance Journal, 1967b, 45, 676-682.

Kaiser, H.F. Little Jiffy, Mark IV. Educational and Psychological Measurement, in press.

Mayo, S.T. Pre-service preparation of teachers in educational measurement. Final Report, U.S.O.E. Project No. 5-0807. Chicago: Loyola University, 1967. (ERIC # SP001516)

Mayo, S.T. Trends in the teaching of the first course in measurement. NCME Symposium Paper, 1970a (ERIC # ED047007)

Mayo, S.T. Measurement in education: mastery learning and mastery testing. NCME Special Report, 1970b, 1(3), 1-4. (ERIC #ED051299)

Schutz, R.E. The role of measurement in education: servant, soulmate, stoolpigeon, statesman, scapegoat, all of the above, and/or none of the above. Journal of Educational Measurement, 1971, 8(3), 141-146.

Stanley, J.C., and Hopkins, K.D. Educational and psychological measurement and evaluation. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Footnotes

[1]Subjects were also asked to rate these objectives for interest. Interest and importance ratings were strongly related. Only importance ratings will be discussed here.

[2]In about half the cases, this was the last five digits of the student's home phone number and, in the remainder, their social security number.

[3]One group of subjects was dropped from the original sample due to non-standard administration of the instrument. Students in early childhood and special education were also excluded from the present analyses.

[4]There were 75 "X" responses (32%) to this item. Otherwise, "X" responses ranged from 0 to 34 with a median close to 5 and a semi-interquartile range of approximately 4.

[5]The Spearman rank order correlation between importance ratings of secondary education sophomores and juniors was .89. Rank order correlations between elementary students' ratings and secondary sophomores' and juniors' ratings were .89 and .90, respectively. (Reliability of the overall rank ordering is also implied in these results.) Multivariate analysis of variance showed no significant differences between the two secondary groups.

[6]A more detailed breakdown follows. Particularly notable are inconsistencies in ranking, where "less important" objectives are strongly related or prerequisite to more highly placed ones.

Three highly rated items involved some aspect of reliability or validity (ranks of 3,6,9), while two other such items (18,28) were further down the list. Evaluation of test usage for improving curriculum & instruction and for feedback and goal setting (5,10) were also very high, while evaluation of usage for placement and grouping and for grading (30,32) were less important. Compared to #32, description of reasons for, and systems of, grading (14) emerged relatively more important.

Uses of objectives were generally rated high with teacher-student definition of objectives, affective objectives, objectives in planning, and different level, and cognitive objectives (2,9,11,13,15) particularly high. Use of behavioral objectives (25) was seen of lesser importance.

Some interesting juxtapositions involved teacher-made tests. Defining a position on when to use teacher-tests in one's classroom (12) was rated fairly high in importance, while various items relating to the actual designing of measures (20,26) including construction of essay and objective items (21,27) were in the middle, or somewhat important, range. Objectives involving item analysis and anecdotal records (33,36) were a little lower. Objectives dealing with semi-objective items and rating scales (40,47) were near the bottom and were clearly less popular than most other types of measures. Of greatest importance was being able to devise new types of instruments where required (4). Being able to contrast different types of measures (23,29), including teacher-made and standardized tests, was seen as somewhat important.

Some other interesting juxtapositions involved standardized tests. Defining a position on one's own usage (8) was seen as very important. However, items relating to specific test-types (22,25,34) were seen as less important as were--and more strongly so--items involving school and district testing programs (38) and standard reference sources (41). Proper administration and scoring (17) was considered fairly important. Appropriate use of norms (19) was rated somewhat high while being able to describe their characteristics and uses (37) was rated much lower.

Defining a position on usage of criterion- and norm-referenced tests in the schools (31) was rated somewhat important, while understanding their current use (39,43) was substantially lower. Four items involving statistics (42,44,45,46)--distributions, central tendency and variability, and correlation--appeared at the bottom of the list. (One item involving correlation (28) was higher where applied to test reliability and validity.) Two other items involved ethical and historical aspects of testing (16,35) with the first seen as fairly important, and relatively more important than the second. The most highly rated item involved making test-taking more enjoyable (1).

[7]Some distributions showed marked skewing. More important here is occasional inhomogeneity of variance, given unequal cell n's (Glass, Peckham, and Sanders, 1972). In particular, alpha values for variables 7 & 8 in the first analysis and variables 1 & 7 in the second analysis, may be higher than reported. Some other effects may be more significant than indicated in this analysis.

[8]An orthoblique solution using Kaiser's (in press) Little Jiffy routine produced no essential improvement; it did, however, produce an overall "index of factorial simplicity" of 0.66 (poor). A direct oblimin procedure was unable to reach convergence in a substantial number of iterations.

[9]The first three factors were particularly strong, relative to the others. Factor I is hard to identify; both its defining and related items involve types of student-oriented test usage (32,30,10), score interpretation and usage (42,43), item/instrument construction (47,27,40), and behaviors related to standardized tests (41,17,24). (Note that while some loadings are relatively low, some communalities are likewise low. Many of the items of extremely low communality showed some restriction of range and might have contributed more powerfully otherwise.)

One defining item also involved objectives in classroom planning (11). The strongest loading item involved test usage in grading (32). Items varied in importance, although many were rated relatively low. Perhaps this factor relates to some global aspect of perceived classroom procedures.

Factor II is defined primarily by several aspects of statistics, all rated relatively low. Factor III involves items, of varying importance, which seem to suggest a broader view of the context in which measurement takes place.

Among the weaker factors, Factor IV is curiously defined by two evaluative items involving uses of tests to improve curriculum and instruction (5) and estimation of instrument validity (6). All relevant items are fairly high in importance. Factor V is loaded on most strongly by making testing more enjoyable (1) and, interestingly, is also defined by construction of essay questions (21), and use of item analysis (33)--objectives of varying importance. Factor VI is related to use of behavioral objectives (25), Factor VII to criterion-referenced tests (31,39), Factor VIII to aspects of reliability and validity (28,9,18), Factors X and XI to special aspects of standardized testing (7,22) and objectives (15,13), and Factor XII is related to devising one's own cognitive and affective measures (26,20).

While these factors cannot be considered to be stable, or the solution unique, this analysis provides some indication of the complex basis on which students are making distinctions between items. Objectives loading on a factor frequently don't show patterns in content, or taxonomy or importance level.

Table 1

Median Importance Ratings of 47 Measurement Objectives for 234 Students in Elementary and Secondary Education

	Objectives*	Median* Ratings	
1.	Devise procedures, make test-taking more enjoyable, my students	1.19	↑ very important →
2.	Devise procedures, teacher-student definition of objectives	1.20	
3.	Distinguish "good" & "bad" tests, usage, testing practice today	1.26	
4.	Devise new types instruments to meet my own needs	1.30	
5.	Evaluate uses of tests to improve curriculum, instruction	1.30	
6.	Decide if instrument measuring what it's supposed to	1.33	
7.	Define own position, standardized test uses, my students	1.39	
8.	Employ educational objectives, affective goals	1.40	
9.	Estimate how consistently a test is doing its job	1.46	
10.	Evaluate uses of tests, inform progress, difficulties, set goals	1.49	
11.	Employ defined educational objectives, planning class activities	1.51	←
12.	Define own position, teacher-made test uses, my classroom	1.52	
13.	Employ educational objectives, different levels complexity	1.56	
14.	Describe reasons for, systems of, grading	1.61	
15.	Employ educational objectives, cognitive goals	1.62	
16.	Define own position, ethical responsibilities of tester	1.69	
17.	Employ appropriate procedures, administering & scoring tests	1.78	
18.	Estimate error in measurements on students	1.80	
19.	Recognize situations where norms useful, misleading, score interpr.	1.81	
20.	Design measures, attainment of cognitive objectives	1.85	
21.	Construct good essay questions	1.85	somewhat important
22.	Discuss features, uses, personality and interest inventories	1.89	
23.	Compare advantages, disadvantages, different types measures	1.87	
24.	Evaluate standardized tests, own subject, grade level taught	1.95	
25.	Employ behavioral objectives	1.97	
26.	Design measures, attainment of affective objectives	2.04	
27.	Construct good objective items	2.05	
28.	Discuss use of correlation in estimating test usefulness	2.09	
29.	Contrast characteristics, standardized and teacher-made tests	2.12	
30.	Evaluate uses of tests for placement and grouping	2.12	
31.	Define own position, use of criterion- & norm-referenced tests	2.14	
32.	Evaluate uses of tests in grading	2.26	
33.	Employ methods of item analysis to improve objective items	2.31	
34.	Compare features, rationales, types intelligence tests	2.32	
35.	Describe origins, testing practice, why testing takes present forms	2.32	
36.	Construct good anecdotal records	2.32	
37.	Describe characteristics, uses, several types norms	2.36	
38.	Outline features, purposes, school and district testing programs	2.47	
39.	Discuss use, criterion-referenced tests	2.61	←
40.	Construct good semi-objective items	2.61	
41.	Locate reviews, information, standardized tests, standard ref. sources	2.63	undecided
42.	Characterize, interpret scores, measures of average, variation	2.73	
43.	Discuss use, norm-referenced tests	2.76	
44.	Use frequency count, histogram, to visualize & interpret scores	2.82	
45.	Examine relationship, two measures, using scatterplot	2.83	
46.	Estimate relationship, two measures, using correlation coefficient	2.88	
47.	Construct good rating scales, behavior checklists	2.94	↓

*Shortened statements

*1 = very important, 2 = somewhat important, 3 = undecided 4 = somewhat unimportant 5 = very unimportant

Table 2

Significant Differences Between Groups on Importance of Measurement Objectives

Item	Group Means* Elem. Fem (N=48)	Sec. Fem. (N=120)	Sec. Males (N=58)	F-Ratio for ANOVA	p<	F-Ratios for Dunn Multiple Contrasts 1 & 2	1 & 3	2 & 3
5. Test uses, impr. C&I	1.83	1.42	1.55	3.74	.03	7.59*		
20. Measure cognitive	2.33	1.91	1.97	3.13	.05	5.96*		
7. Own use stand. tests	1.58	1.56	1.93	3.21	.04			5.94*
8. Affective objectives	1.65	1.60	2.00	3.68	.03			6.94*
17. Admin., scoring	2.00	1.83	2.41	5.47	.005			10.74**
22. Personality, interest	1.94	1.88	2.34	5.08	.01			9.54**
27. Objective Items	2.19	2.21	2.76	4.09	.02			7.04*
28. Correl'n, test value	1.96	2.17	2.60	6.33	.002		11.38**	7.51*
31. Crit. & norm-ref. tests	2.06	2.19	2.57	3.11	.05		--	--
45. Scatterplot	2.71	2.82	3.24	4.16	.02		6.60*	6.06*
46. Correlation, two meas.	2.83	2.82	3.24	6.80	.001		8.00*	12.18**

*p<.05
**p<.01

Item	H/SS (N=77)	M/S (N=41)	F-Ratio Main Effect on Specialty	p<
27. Objective items	2.62	2.05	6.57	.01
32. Test uses, grading	2.70	2.15	5.75	.02
1. Testing enjoyable	1.34	1.73	5.76	.02
7. Own use, stand. tests	1.61	2.05	5.55	.02
13. Diff. levels, objectives	1.68	2.02	4.03	.05
19. Where norms useful	1.74	2.42	18.51	.0001
21. Essay questions	1.94	2.93	18.01	.0001
22. Personality, interest	1.94	2.39	6.04	.02
Stand., tchr. tests	2.17	2.66	5.33	.02
Origins, testing	2.39	3.10	7.98	.006
37. Types, uses, norms	2.40	2.93	6.42	.01
38. District testing	2.66	3.15	5.18	.02
45. Scatterplot	2.82	3.29	5.89	.02

* Highest value is underlined for each variable. (Larger numbers correspond to relatively lower ratings)

Table 3

Varimax Factor Solution for Importance Ratings of 47 Measurement Objectives (N=234)

Items	Factor Loadings													h
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	
32. Test uses, grading	71													6
42. Average, variation	59													4
11. Objectives, planning	56													4
47. Rating scales	56													4
41. References, tests	37													3
27. Objective items	51												35	5
30. Test uses, plcmt, gpg.	49		38											4
40. Semi-objective items	48												41	5
43. Use, norm-ref. tests	46								36					4
17. Admin., scoring	44			31								41		6
24. Stand. tests, own subj.	38	32												4
10. Test uses, progress	35			32										2
45. Scatterplot		63												5
44. Histogram		51												4
46. Correl'n two meas.		47												3
37. Types, uses norms		39												4
29. Stand., tchr.- tests		49	31											5
34. Intelligence tests		35					33							3
35. Origins testing			60											5
14. Systems grading			57											4
16. Ethical responsibility			56											4
2. Tchr-stu. objectives			36											3
38. District testing		38	47											5
3. "Good", "bad" testing			43					37						4
5. Test uses, impr. C&I				66										5
6. Meas. what supp. to				48										5
4. New types measures			31	40					32	33				5
13. Adv., disadv., diff. meas.		32	37	37										5
1. Testing enjoyable					67									6
21. Essay questions					53									4
33. Item analysis					46									5
25. Behavioral objectives						65								4
12. Own use tchr.-tests						37		31						4
36. Anecdotal records		(20)				(26)								2
31. Crit. & norm-ref. tests							72							6
39. Uses crit-ref. tests							54							5
28. Correl'n, test value								55						5
9. Test, consistent job								51						4
18. Estimate error								37						5
7. Own use, stand. test									65					6
22. Personality, interest									42					4
15. Cognitive objectives										58				4
13. Diff. levels, objectives						49				53				6
16. Measure affective											54			4
20. Measure cognitive											41			4
18. Affective objectives					34	34					35			5
9. Where norms useful		38	36									42		5

Note.--Factor loadings ≥ .30 are included in the table. Decimal points have been omitted.